

## **RAW SEQUENCE LISTING**

**The Biotechnology Systems Branch of the Scientific and Technical  
Information Center (STIC) no errors detected.**

Application Serial Number: 10/549,943  
Source: IFWO  
Date Processed by STIC: 04/05/2007

# ***ENTERED***



IFWO

## RAW SEQUENCE LISTING

DATE: 04/05/2007

PATENT APPLICATION: US/10/549,943

TIME: 07:50:22

Input Set : A:\796\_2\_PCT\_SeqListing.TXT

Output Set: N:\CRF4\04052007\J549943.raw

```

4 <110> APPLICANT: Genencor International, Inc.
5       Jones, Brian E.
6       Grant, William D.
7       Heaphy, Shaun
8       Grant, Susan
10 <120> TITLE OF INVENTION: Novel Bacillus 029cel Cellulase
13 <130> FILE REFERENCE: GC796-2-PCT
C--> 15 <140> CURRENT APPLICATION NUMBER: US/10/549,943
C--> 16 <141> CURRENT FILING DATE: 2005-09-20
18 <150> PRIOR APPLICATION NUMBER: US 60/466,831
19 <151> PRIOR FILING DATE: 2003-04-29
21 <160> NUMBER OF SEQ ID NOS: 3
23 <170> SOFTWARE: FastSEQ for Windows Version 4.0
25 <210> SEQ ID NO: 1
26 <211> LENGTH: 3410
27 <212> TYPE: DNA
28 <213> ORGANISM: Bacillus sp.
30 <220> FEATURE:
31 <221> NAME/KEY: misc_feature
32 <222> LOCATION: (1)...(3410)
33 <223> OTHER INFORMATION: isolated from environmental sample from Sonachi Lake, Kenya
35 <400> SEQUENCE: 1
36 atcaacacgc tggaaagtaa tttcaagggt aaggccatcg gttgccgcgc gggtagaaat      60
37 gtgcggttg atttcgttga gcggcgtcgc cggcggtcca ccgagggcat agcgcagcag      120
38 gttggcgatg ccaccggtga ggccttcggg gccgcctacg atgttggtgct cagccgccca      180
39 tgcgatgtag ccgtccggct cgggttcgct cgcgggggtg aagaagacaa tgcgtcgag      240
40 ataaagggtg ccgcttcgcg tctcaacgcc gccgaggttg aattggattt cgcaaattct      300
41 cgttagggtc agcacggaat cgcgcacgag gtcggctatg ggaatctgaa tgcgcccata      360
42 ggggttggtg cgcggaaggg acacgtaggg acccactttg tcattgggag agacgagccg      420
43 gacaaagatt tgggtgcgccg cctgcgaggg gccttgaggg gcgagagaaa ggtacgtgag      480
44 ggcgctgatg tcgtgcgtgg gaccgtctcc ccagttgtcg agattgagcc caaatccggc      540
45 ccaccatccg gcgatagtgt agtccaatg gtagtgacgc tcaccctcga agccgcgct      600
46 ggagagttcc tgcaagccgt cgcgccaaat gcccgatg agcgttgctt cgtcacggta      660
47 gatcacaagt tcggcggcgg gtgccggggg aagatcgctt tgagtgatca cgagagtggc      720
48 ggtggcgctg ccttcgtgat tagggctcgg aatgggtggc acgaccgtgt agctaccggg      780
49 cccactggc gcattgggtg aaccgttgta ggtaaaggag acgtcaagcc ccacgggatg      840
50 ggtctcggca agagcggcct tgggggtgcc gtcgaaaacg tgttccaaat tggagagcgt      900
51 gatggtggcg ggtgccttga gcacagtcac agaaacagtg gattgcacgg gatcgtgcgc      960
52 tgccgtgtct gcaggtgtga agaccacgct gtaaaaacgg gttccggcgg acggtgcaag      1020
53 gccggacagg acaaaggcaa agtcgcgggg gacggcggtc actccgccgc tcaggccggc      1080
54 ctccgcaagg gtttgcgcga aggtgatggg tgcggctgtg ggccacatct ccacaaggcc      1140
55 ggtgtcccc tcgtcacgca ccggcatgag ggcgagagg agatgaatgt aactggcttg      1200
56 gtaattgatg tcgggctcgg tgatttccca tgagttctcc ggccaaaaac cattccaatc      1260

```

## RAW SEQUENCE LISTING

DATE: 04/05/2007

PATENT APPLICATION: US/10/549,943

TIME: 07:50:22

Input Set: A:\796\_2\_PCT\_SeqListing.TXT

Output Set: N:\CRF4\04052007\J549943.raw

```

57 aaggtaggct ttttgcacgg gttgggtctcg gatcgctga atgcttccgc tgtatttggg 1320
58 cattgggacc cgcccgaag aaaaccagga gcgggaccgt agagtgaagt gagggcattg 1380
59 tcccagtcgg gccatcgcg aaccaatggg ggtagatttc attggctgca cggtcagcgc 1440
60 cgctggcata catgttgcta agatagacca tgcccattgg gttcactccg tggagatagt 1500
61 gcaggtagcc catcgcgga tcgcatgctg cggccgcgct ggcgggggtg agcccaagcc 1560
62 tccgtacccc ctgaagaaa aagccagcct gagactttgt tttgttcgag cccacgtgt 1620
63 aatcctgatc cttcaggtag gcgcggtagg cgtcgggtctg gttattccat gcaccgagaa 1680
64 actccccacc gtttatagaa gccgccatcc ggttgcggtat gtcggcagag acgctaggcg 1740
65 tcgctcccgg gagggctcgt tagtgggcca gagctttttg tagctcacct tgaaagggga 1800
66 agaaatacca cactgcacg ggctccatat cgagatagcg cacatcgaag aaatcgcgat 1860
67 agaccgcacc gcccgctgcg tcgaagagca tggcgggcgc catcacacgg ttggctagcg 1920
68 tatcgtgggc attgcgcgag gggctcacgg aagcaaattc ggtgttgctg aaaggcacat 1980
69 gaggatggac catggtccaa ttccatgcgg cgatggcagc ggattcgagg gtgacggcat 2040
70 aatcgctcat gcctacgctc tcaaagacag tcgccccgag ggcgaaagcg gcggcagcca 2100
71 tggcagtggc ctcggtcgag acggggccgt agtaacgcgg atgggtgtcg gtgctcggcg 2160
72 ggctggcgct ctggtgcccc gtcacggaaa ctttcccgag aatagccccg ctcggtcct 2220
73 gcatgcgtaa gagccagtcc attccccatt tgacttcgct aagcaggctc gggacaccgt 2280
74 tgccggattc cgggatgcc aatcatcggt taaagacgct aggccgccct tgataggcaa 2340
75 ggagcagctc caggatgacg cggcccgctc actcgctgta cttgttgaaa tcgcccgcct 2400
76 cgaaccaacc gccgctgaga tcgctgctca aggaggcatt ccccatatcc cagatggggc 2460
77 ggctggcgac gtctgctggg tgagaagcgg catcggccca gttcgctggg gcgtagggca 2520
78 cctccttggc aaaccgggag cgctgataga agaactgctg cacggcctcg cgcaggacaa 2580
79 catcgtaaac atccgcgcca atggcgaaac tatcggaatg agtgttggtg gcaggatcgt 2640
80 ggatgcggta gtggccgggc tcggcaacta ccgtaaaatc aaaccaccac acgcggtctc 2700
81 ccgattgaat atggatggcg ccgcccgttc acgggaccgg tgagccggag aaaaccacga 2760
82 cgccatcggt cagcgacgg acctccagcg ttgcgcccgg gctgtagctc tcggcgctgt 2820
83 tccagccaat ctgcgggtcg gcgatcaccg ccaccttggg ggcacggcg gggtaaccga 2880
84 attggtcgat gcggatttta tcggtgtggg tggaggcgac gagggcgagg ctgcccata 2940
85 gcagcaagaa aaagcccgtc gtcggcccga taccaaaaaa acgaataggg agagaaaaat 3000
86 tcatagcagg atgtggatac ggaaagggg aaaacggctg aaagaccaa gcccaacgt 3060
87 tggcgaaaac tggatggtt gtttatcaag aaaagcgctt ttgagccaaa agctgcgggc 3120
88 aatccttatt gcgtttcaca atattttcac atcgctcggc gcacgacttt tcgatgggcg 3180
89 acttgacagc gtattctctc aggcgcgagg ctgcaaacct tatgaaaaaa ggcccgcgca 3240
90 gcgatctgtc cccggtcaaa atccagtcaa ggtttgttca agggtttgag gtctgataga 3300
91 ggcacagtgc agccatcagc agtcgcattg agtaggggtg ttggagaaag tgtgcaaatg 3360
92 accgctgccc aaggaactgt ggagacaaaa agcatatttt cctcgccaag 3410

```

94 &lt;210&gt; SEQ ID NO: 2

95 &lt;211&gt; LENGTH: 1746

96 &lt;212&gt; TYPE: DNA

97 &lt;213&gt; ORGANISM: Bacillus sp.

99 &lt;220&gt; FEATURE:

100 &lt;221&gt; NAME/KEY: misc\_feature

101 &lt;222&gt; LOCATION: (1)...(1746)

102 &lt;223&gt; OTHER INFORMATION: isolated from environmental sample from Sonachi Lake, Kenya

104 &lt;400&gt; SEQUENCE: 2

```

105 atgaattttt ctctccctat tcgttttttt ggtatcgggc cgacagcggg ctttttcttg 60
106 ctgctcatgg gcagctccgc cctcgctgcc tccaccaca ccgataaaat ccgcatcgac 120
107 caattcgggt accccgccga tgccaccaag gtggcggtga tcgccgacc gcagattggc 180
108 tggaacagcg ccgagagcta cagccccgc gcaacgctgg aggtccgtcg cgtgaacgat 240

```

## RAW SEQUENCE LISTING

DATE: 04/05/2007

PATENT APPLICATION: US/10/549,943

TIME: 07:50:22

Input Set : A:\796\_2\_PCT\_SeqListing.TXT

Output Set: N:\CRF4\04052007\J549943.raw

```

109 ggcgtcgtgg ttttctccgg ctccaccggtc ccgtggaacg gcggcgccat ccatattcaa 300
110 tcgggagacc gcgtgtggtg gtttgatttt acggtagtgt ccgagcccgg ccactaccgc 360
111 atccacgata ctgccaacaa cactcattcc gatagtgttc ccattggcgc ggatgtttac 420
112 gatgttgtcc tgcgcgaggg cgtgcgcgat ttcttctatc agcgtccgg gtttgccaag 480
113 gaggtgccct acgcccacgc gaactggggc gatgccgctt ctccccgca ggacgtcgcc 540
114 agccgcccc a tctgggatat ggggaatgcc tccttgaggc gcgatctcag cggcggttg 600
115 ttcgatgcgg gcgatttcaa caagtacagc gaggggacgg ggcgcgtcat cctggagctg 660
116 ctcttgccct atcaagggcg gcctgacgtc tttaccgatg attttggcat cccggaatcc 720
117 ggcaacgggtg tccccgacct gcttgacgaa gtcaaatggg gaatggactg gctcttacgc 780
118 atgcaggagc cgagcggggc tattctcggg aaagtttccg tgacggggca ccagagcgcc 840
119 agcccgcga gcaccgacac ccattccgct tactacggcc ccgtctcgac cgaggccact 900
120 gccatggctg ccgcccgttt cgccctcggg gcgactgtct ttgagagcgt aggcattgagc 960
121 gattatgccg tcaccctcga atccgtgcc atcgccgat ggaattggac catggtccat 1020
122 cctcatgtgc ctttcgacaa caccggattt gcttcggtga gcccctcgcg caatgcccac 1080
123 gatacgctag ccaaccgtgt gatggcgccc gccatgctct tcgagcgcac gggcggtgcg 1140
124 gtctatcgcg atttcttcga tgtgcgtat ctcgatatgg agcccgcgca gtggtggtat 1200
125 ttcttcccc ttcaaggtga gctacaaaaa gctctcgccc actacacgac cctcccgga 1260
126 gcgacgecta gcgtctctgc cgacatccgc aaccggatgg cggcttctat aaacggtggg 1320
127 gaqtttctcg gtgcatggaa taaccagacc gacgcctacc gcgcctacct gaaggatcag 1380
128 gattacacgt ggggctcgaa caaaacaaag tctcaggctg gctttttctt cgagggggta 1440
129 cggaggcttg ggctcaaccc cgccgacgcg gccgcgcata gcgatgccgc gatgggctac 1500
130 ctgcactatc tccacggagt gaacccaatg ggcattggtct atcttagcaa catgtatgcc 1560
131 agcggcgctg accgtgcagc caatgaaatc taccaccatt ggttcgcgca tggccggact 1620
132 gggacaatgc cctcacttca ctctacggtc ccgtcctggt ttttctttcg ggcgggtccc 1680
133 aatgccc aaa tacagcgga gcatcaggc gatccgagac caaccggtgc aaaaagccta 1740
134 cttga

```

136 &lt;210&gt; SEQ ID NO: 3

137 &lt;211&gt; LENGTH: 581

138 &lt;212&gt; TYPE: PRT

139 &lt;213&gt; ORGANISM: Bacillus sp.

141 &lt;220&gt; FEATURE:

142 &lt;221&gt; NAME/KEY: VARIANT

143 &lt;222&gt; LOCATION: (1)...(581)

144 &lt;223&gt; OTHER INFORMATION: isolated from environmental sample from Sonachi Lake, Kenya

146 &lt;400&gt; SEQUENCE: 3

147 Met Asn Phe Ser Leu Pro Ile Arg Phe Phe Gly Ile Gly Pro Thr Ala

148 1 5 10 15

149 Gly Phe Phe Leu Leu Leu Met Gly Ser Ser Ala Leu Val Ala Ser Thr

150 20 25 30

151 His Thr Asp Lys Ile Arg Ile Asp Gln Phe Gly Tyr Pro Ala Asp Ala

152 35 40 45

153 Thr Lys Val Ala Val Ile Ala Asp Pro Gln Ile Gly Trp Asn Ser Ala

154 50 55 60

155 Glu Ser Tyr Ser Pro Gly Ala Thr Leu Glu Val Arg Arg Val Asn Asp

156 65 70 75 80

157 Gly Val Val Val Phe Ser Gly Ser Pro Val Pro Trp Asn Gly Gly Ala

158 85 90 95

159 Ile His Ile Gln Ser Gly Asp Arg Val Trp Trp Phe Asp Phe Thr Val

160 100 105 110

## RAW SEQUENCE LISTING

DATE: 04/05/2007

PATENT APPLICATION: US/10/549,943

TIME: 07:50:22

Input Set : A:\796\_2\_PCT\_SeqListing.TXT

Output Set: N:\CRF4\04052007\J549943.raw

```

161 Val Ala Glu Pro Gly His Tyr Arg Ile His Asp Pro Ala Asn Asn Thr
162      115      120      125
163 His Ser Asp Ser Phe Ala Ile Gly Ala Asp Val Tyr Asp Val Val Leu
164      130      135      140
165 Arg Glu Ala Val Arg Met Phe Phe Tyr Gln Arg Ser Gly Phe Ala Lys
166 145      150      155      160
167 Glu Val Pro Tyr Ala His Ala Asn Trp Ala Asp Ala Ala Ser His Pro
168      165      170      175
169 Gln Asp Val Ala Ser Arg Pro Ile Trp Asp Met Gly Asn Ala Ser Leu
170      180      185      190
171 Glu Arg Asp Leu Ser Gly Gly Trp Phe Asp Ala Gly Asp Phe Asn Lys
172      195      200      205
173 Tyr Ser Glu Trp Thr Gly Arg Val Ile Leu Glu Leu Leu Ala Tyr
174      210      215      220
175 Gln Gly Arg Pro Asp Val Phe Thr Asp Asp Phe Gly Ile Pro Glu Ser
176 225      230      235      240
177 Gly Asn Gly Val Pro Asp Leu Leu Asp Glu Val Lys Trp Gly Met Asp
178      245      250      255
179 Trp Leu Leu Arg Met Gln Glu Pro Ser Gly Ala Ile Leu Gly Lys Val
180      260      265      270
181 Ser Val Thr Gly His Gln Ser Ala Ser Pro Pro Ser Thr Asp Thr His
182      275      280      285
183 Pro Arg Tyr Tyr Gly Pro Val Ser Thr Glu Ala Thr Ala Met Ala Ala
184      290      295      300
185 Ala Ala Phe Ala Leu Gly Ala Thr Val Phe Glu Ser Val Gly Met Ser
186 305      310      315      320
187 Asp Tyr Ala Val Thr Leu Glu Ser Ala Ala Ile Ala Ala Trp Asn Trp
188      325      330      335
189 Thr Met Val His Pro His Val Pro Phe Asp Asn Thr Gly Phe Ala Ser
190      340      345      350
191 Val Ser Pro Ser Arg Asn Ala His Asp Thr Leu Ala Asn Arg Val Met
192      355      360      365
193 Ala Ala Ala Met Leu Phe Glu Arg Thr Gly Gly Ala Val Tyr Arg Asp
194      370      375      380
195 Phe Phe Asp Val Arg Tyr Leu Asp Met Glu Pro Val Gln Trp Trp Tyr
196 385      390      395      400
197 Phe Phe Pro Phe Gln Gly Glu Leu Gln Lys Ala Leu Ala His Tyr Thr
198      405      410      415
199 Thr Leu Pro Gly Ala Thr Pro Ser Val Ser Ala Asp Ile Arg Asn Arg
200      420      425      430
201 Met Ala Ala Ser Ile Asn Gly Gly Glu Phe Leu Gly Ala Trp Asn Asn
202      435      440      445
203 Gln Thr Asp Ala Tyr Arg Ala Tyr Leu Lys Asp Gln Asp Tyr Thr Trp
204      450      455      460
205 Gly Ser Asn Lys Thr Lys Ser Gln Ala Gly Phe Phe Phe Glu Gly Val
206 465      470      475      480
207 Arg Arg Leu Gly Leu Asn Pro Ala Asp Ala Ala Ala His Arg Asp Ala
208      485      490      495
209 Ala Met Gly Tyr Leu His Tyr Leu His Gly Val Asn Pro Met Gly Met

```

## RAW SEQUENCE LISTING

DATE: 04/05/2007

PATENT APPLICATION: US/10/549,943

TIME: 07:50:22

Input Set : A:\796\_2\_PCT-SeqListing.TXT

Output Set: N:\CRF4\04052007\J549943.raw

210				500				505					510			
211	Val	Tyr	Leu	Ser	Asn	Met	Tyr	Ala	Ser	Gly	Ala	Asp	Arg	Ala	Ala	Asn
212				515				520						525		
213	Glu	Ile	Tyr	His	His	Trp	Phe	Arg	Asp	Gly	Arg	Thr	Gly	Thr	Met	Pro
214			530				535						540			
215	Ser	Leu	His	Ser	Thr	Val	Pro	Leu	Leu	Val	Phe	Phe	Arg	Ala	Gly	Pro
216	545					550					555				560	
217	Asn	Ala	Gln	Ile	Gln	Arg	Lys	His	Ser	Gly	Asp	Pro	Arg	Pro	Thr	Arg
218					565					570					575	
219	Ala	Lys	Ser	Leu	Pro											
220					580											

VERIFICATION SUMMARY

DATE: 04/05/2007

PATENT APPLICATION: US/10/549,943

TIME: 07:50:23

Input Set : A:\796\_2\_PCT\_SeqListing.TXT

Output Set: N:\CRF4\04052007\J549943.raw

L:15 M:270 C: Current Application Number differs, Replaced Current Application Number

L:16 M:271 C: Current Filing Date differs, Replaced Current Filing Date